DigitalCommons@WayneState

Journal of Modern Applied Statistical Methods

Volume 4 | Issue 1 Article 5

5-1-2005

Coverage Properties Of Optimized Confidence Intervals For Proportions

John P. Wendell
University of Hawai`i at Mānoa

Sharon P. Cox
University of Hawai`i at Mānoa

Journal of Modern Applied Statistical Methods
May, 2005, Vol. 4, No.1, 43-52

1538 – 9472/05/$95.00

Coverage Properties Of Optimized Confidence Intervals For Proportions

John P. Wendell Sharon P. Cox
College of Business Administration
University of Hawai`i at Mānoa

Wardell (1997) provided a method for constructing confidence intervals on a proportion that modifies the Clopper-Pearson (1934) interval by allowing for the upper and lower binomial tail probabilities to be set in a way that minimizes the interval width. This article investigates the coverage properties of these optimized intervals. It is found that the optimized intervals fail to provide coverage at or above the nominal rate over some portions of the binomial parameter space but may be useful as an approximate method.

Key words: Attribute, Bernoulli, dichotomous, exact, sampling

Introduction

A common task in statistics is to form a confidence interval on the binomial proportion *p*. The binomial probability distribution function is defined as

$$\Pr\left[Y = y \mid p, n\right] = \mathrm{b}\left(p, n, y\right)$$
$$= \binom{n}{y} p^{y}\left(1 - p^{n-y}\right),$$

where the proportion of elements with a specified characteristic in the population is *p*, the sample size is *n*, and *y* is the outcome of the random variable *Y* representing the number of elements with a specified characteristic in the sample.

The coverage probability for a given value of *p* is

$$C_{n,CL^*}(p) = \sum_{i=0}^{n} \mathrm{I}(i, p)\mathrm{b}(p, n, i),$$

where $C_{n,CL^*}(p)$ is the coverage probability for a particular method with a nominal confidence level CL^* for samples of size *n* taken from a population with binomial parameter *p* and $\mathrm{I}(i, p)$ is 1 if the interval contains *p* when $y = i$ and 0 otherwise. The actual confidence level of a method for a given CL^* and *n* $\left(CL_{n,CL^*}\right)$ is the infimum over *p* of $C_{n,CL^*}(p)$. Exact confidence interval methods (Blyth & Still, 1983) have the property that $CL_{n,CL^*} \geq CL^*$ for all *n*, and CL^*.

The most commonly used exact method is due to Clopper and Pearson (1934) and is based on inverting binomial tests of $H_0 : p = p_0$. The upper bound of the Clopper-Pearson interval (*U*) is the solution in p_0 to the equation

$$\sum_{i=y}^{n} \mathrm{b}\left(\mathrm{p}_0, n, i\right) = \alpha_U ,$$

except that when $y = n$, $U = 1$. The lower bound, *L*, is the solution in p_0 to the equation

$$\sum_{i=0}^{y} \mathrm{b}\left(\mathrm{p}_0, n, i\right) = \alpha_L ,$$

except that when $y = 0$, $L = 0$. The nominal confidence level $CL^* = 1 - \alpha$ where

John P. Wendell is Professor, College of Business Administration, University of Hawai`i at Mānoa. E-mail: cbaajwe@hawaii.edu. Sharon Cox is Assistant Professor, College of Business Administration, University of Hawai`i at Mānoa.

$\alpha = \alpha_U + \alpha_L$. Because the Clopper-Pearson bounds are determined by inverting hypothesis tests, both α_U and α_L are set *a priori* and remain fixed regardless of the value of *y*. In practice, the values of α_U and α_L are often set to $\alpha_U = \alpha_L = \alpha / 2$.

Wardell (1997) modified the Clopper-Pearson bounds by replacing the condition that α_U and α_L are fixed with the condition that only α is fixed. This allows α to be partitioned differently between α_U and α_L for each sample outcome *y*. Wardell (1997) provided an algorithm for accomplishing this partitioning in such a way that the confidence interval width is minimized for each *y*. Intervals calculated in this way are referred to here as optimized intervals. Wardell (1997) was concerned with determining the optimized intervals and not the coverage properties of the method. The purpose of this article is to investigate the coverage properties.

Coverage Properties of Optimized Intervals

Figure 1 plots $C_{n,.95}(p)$ against *p* for sample sizes of 5, 10, 20, and 50. The discontinuity evident in the Figure 1 plots is due to the abrupt change in the coverage probability when *p* is at *U* or *L* for any of the $n+1$ confidence intervals. Berger and Coutant (2001) demonstrated that the optimized interval method is an approximate and not an exact method by showing that $CL_{5,.95} = .9375 < .95$. Figure 1 confirms the Berger and Coutant result and extends it to sample sizes of 10, 20, and 50.

Agresti and Coull (1998) argued that some approximate methods have advantages over exact methods that make them preferable in many applications. In particular, they recommended two approximate methods for use by practitioners: the score method and adjusted Wald method. The interval bounds for the score method are

$$\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\left[\hat{p}\left(1-\hat{p}\right) + z_{\alpha/2}^2 / 4n \right] / n} \right) / \left(1 + z_{\alpha/2}^2 / n\right),$$

where $\hat{p} = y / n$ and z_c is the $1-c$ quantile of the standard normal distribution. The adjusted Wald method interval bounds are

$$\tilde{p} \pm z_{\alpha/2} \sqrt{\tilde{p}\left(1-\tilde{p}\right) / \left(n+4\right)},$$

where

$$\tilde{p} = \left(y+2\right) / \left(n+4\right).$$

One measure of the usefulness of an approximate method is the average coverage probability over the parameter space when *p* has a uniform distribution. This measure is used by Agresti and Coull (1998). Ideally, the average coverage probability should equal the nominal coverage probability. Figure 2 is a plot of the average coverage probabilities for the optimized interval, adjusted Wald and score methods for sample sizes of 1 to 100 and nominal confidence levels of .80, .90, 95, and .99.

Both the adjusted Wald and the score method perform better on this measure than the optimized interval method in the sense the average coverage probability is closer to the nominal across all of the nominal confidence levels and sample sizes. However, the optimized interval method has the desirable property that the average coverage probability never falls below the nominal for any of the points plotted. The score method is below the nominal for the entire range of sample sizes at the nominal confidence level of .99 and the same is true for the adjusted Wald method at the nominal confidence level of .80.

Figure 1. Coverage Probabilities of Optimized Intervals Across Binomial Parameter p. The disjointed lines plot the actual coverage probabilities of the optimized interval method across the entire range of values of p at a nominal confidence level of .95 for sample sizes of 5, 10, 20, and 50. The discontinuities occur at the boundary points of the n + 1 confidence intervals. The horizontal dotted line is at the nominal confidence level of .95. For all four sample sizes the actual coverage probability falls below the nominal for some values of p, demonstrating that the optimized bounds method is not an exact method.

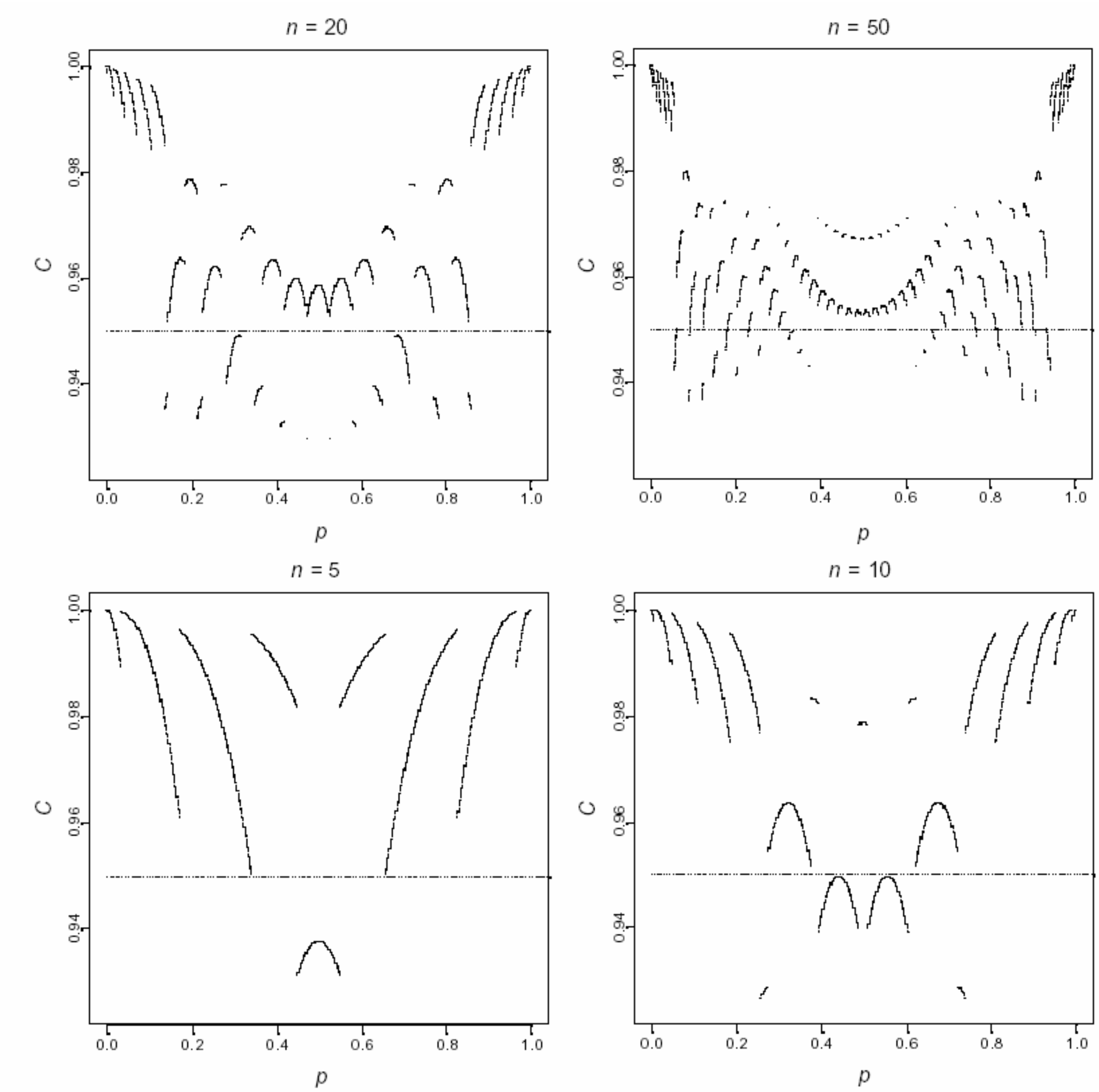
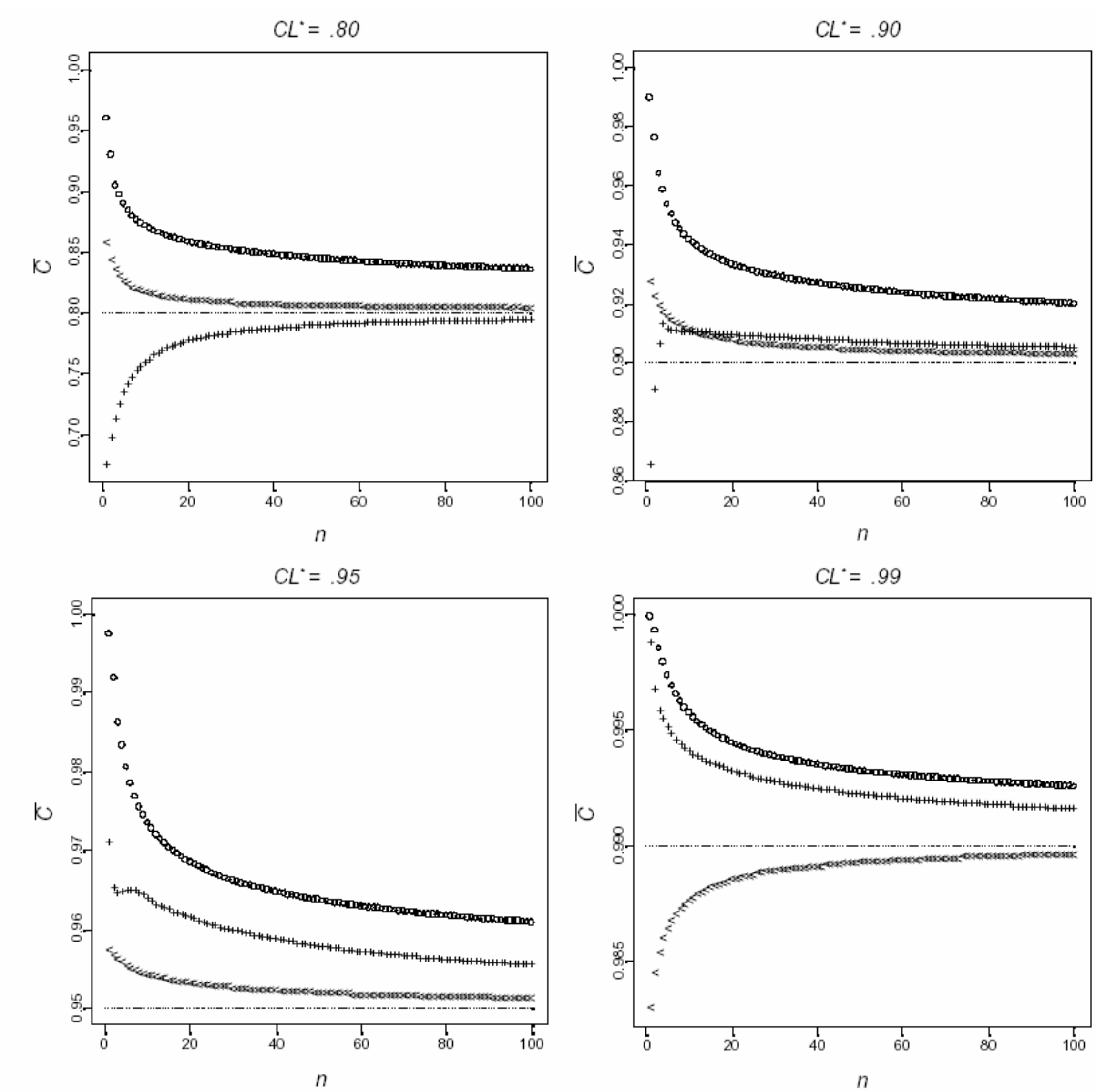


Figure 2. Average Coverage Probabilities of Three Approximate Methods. The scatter is of the average coverage probabilities of three approximate methods when p is uniformly distributed for sample sizes of from 1 to 100 with nominal confidence levels of .80, .90, .95, and .99. The optimized interval method is indicated by a “o”, the adjusted Wald method by a “+”, and the score method by a “<”. The horizontal dotted line is at the nominal confidence level. The optimized interval method’s average coverage probability tends to be further away from the nominal than the other two methods for all four nominal confidence levels and is always higher than the nominal. The average coverage probabilities of the other two methods tend to be closer to, and sometimes below, the nominal level.



A second measure used by Agresti and Coull (1998) is $\sqrt{\int_0^1 \left(C_{n,CL^*}(p) - CL^*\right)^2 dp}$, the uniform-weighted root mean squared error of the average coverage probabilities about the nominal confidence level. Ideally, this mean squared error would equal zero. Figure 3 plots the root mean squared error for the three methods over the same range of sample sizes and nominal confidence levels as Figure 2. The relative performance of the three methods for this metric varies according to the nominal confidence level. Each method has at least one nominal confidence level where the root mean squared error is furthest from zero for most of the sample sizes. The score method is worst at nominal confidence level of .99, the adjusted Wald at .80, and the optimized interval method at both .90 and .95.

Agresti and Coull (1998) also advocated comparing one method directly to another by measuring the proportion of the parameter space where the coverage probability is closer to the nominal for one method than the other. Figure 4 plots this metric for both the score method and the adjusted Wald method versus the optimized interval method for the same sample sizes and nominal confidence levels as Figures 2 and 3.

The results are mixed. At the .99 nominal confidence level the coverage of the adjusted Wald method is closer to the nominal in less than 50% of the range of *p* for all sample sizes, whereas the score method is closer for more than 50% of the range of *p* for all sample sizes above 40. At the other three nominal confidence levels both the adjusted Wald and score methods are usually closer to the nominal than the optimized interval method in more 50% of the range of *p* when sample sizes are greater than 20 and less than 50% for smaller sample sizes. Neither method is closer than the optimized interval method to the nominal confidence level in more than 65% of the range of *p* for any of the pairs of sample sizes and nominal confidence levels.

Another metric of interest is the proportion of the range of *p* where the coverage probability is less than the nominal. For exact methods, this proportion is zero by definition. For approximate methods, a small proportion of the range of *p* with coverage probabilities less than the nominal level is preferred. Figure 5 plots this metric over the same sample sizes and nominal confidence levels as Figures 2 to 4. The optimized interval method is closer to zero than the other methods for almost all of the sample sizes and nominal confidence levels. The adjusted Wald is the next best, with the score method performing the worst on this metric.

The approximate methods all have the property that $CL_{n,CL^*} < CL^*$ for most values of CL^* and *n*, so it is of interest how far below the nominal confidence level the actual confidence level is. The actual coverage probability of the optimized interval method can never fall below the nominal minus α, that is $CL_{n,CL^*} \geq CL^* - \alpha$ for every *n* and CL^*. This follows from the restriction that $\alpha_U + \alpha_L = \alpha$ which requires that α_U and $\alpha_L \leq \alpha$ for all *y*. As a result, the $CL^* = 1 - \alpha$ level optimized intervals must be contained within the Clopper-Pearson $CL^* = 1 - 2\alpha$ level intervals. Because the Clopper-Pearson method is an exact method, it follows directly that $CL_{n,CL^*} \geq CL^* - \alpha$ for all *n* and CL^*. The score and the adjusted Wald method have no such restriction on CL_{n,CL^*}.

Figure 6 plots the actual coverage probability of the optimized interval method against sample sizes ranging from 1 to 100 for nominal confidence levels of .80, .90, .95, and .99. Figure 6 shows that the optimized method is always below the nominal except for very small sample sizes. It is often within a distance of $\alpha/2$ of the nominal confidence level, particularly for sample sizes over 20. The performance of the adjusted Wald method for this metric is very similar to the optimized interval method for sample sizes over 10 at the .95 and .99 confidence level. At the .80 and .90 confidence level the adjusted Wald performs very badly, with coverage probabilities of zero for all of the sample sizes when the nominal level is .80. The score method is the opposite, with actual confidence levels substantially below the nominal at the .95 and .99 nominal levels and closer at the .90 and .80 levels.

Figure 3. Root Mean Square Error of Three Approximate Methods. The scatter is of the uniform-weighted root mean squared error of the average coverage probabilities of three approximate methods when p is uniformly distributed for sample sizes of from 1 to 100 with nominal confidence levels of .80, .90, .95, and .99. The optimized interval method is indicated by a “o”, the adjusted Wald method by a “+”, and the score method by a “<”. The relative performance of the three methods for this metric varies according to the nominal confidence level. Each method has at least one nominal confidence level where the root mean squared error is furthest from zero for most of the sample sizes.

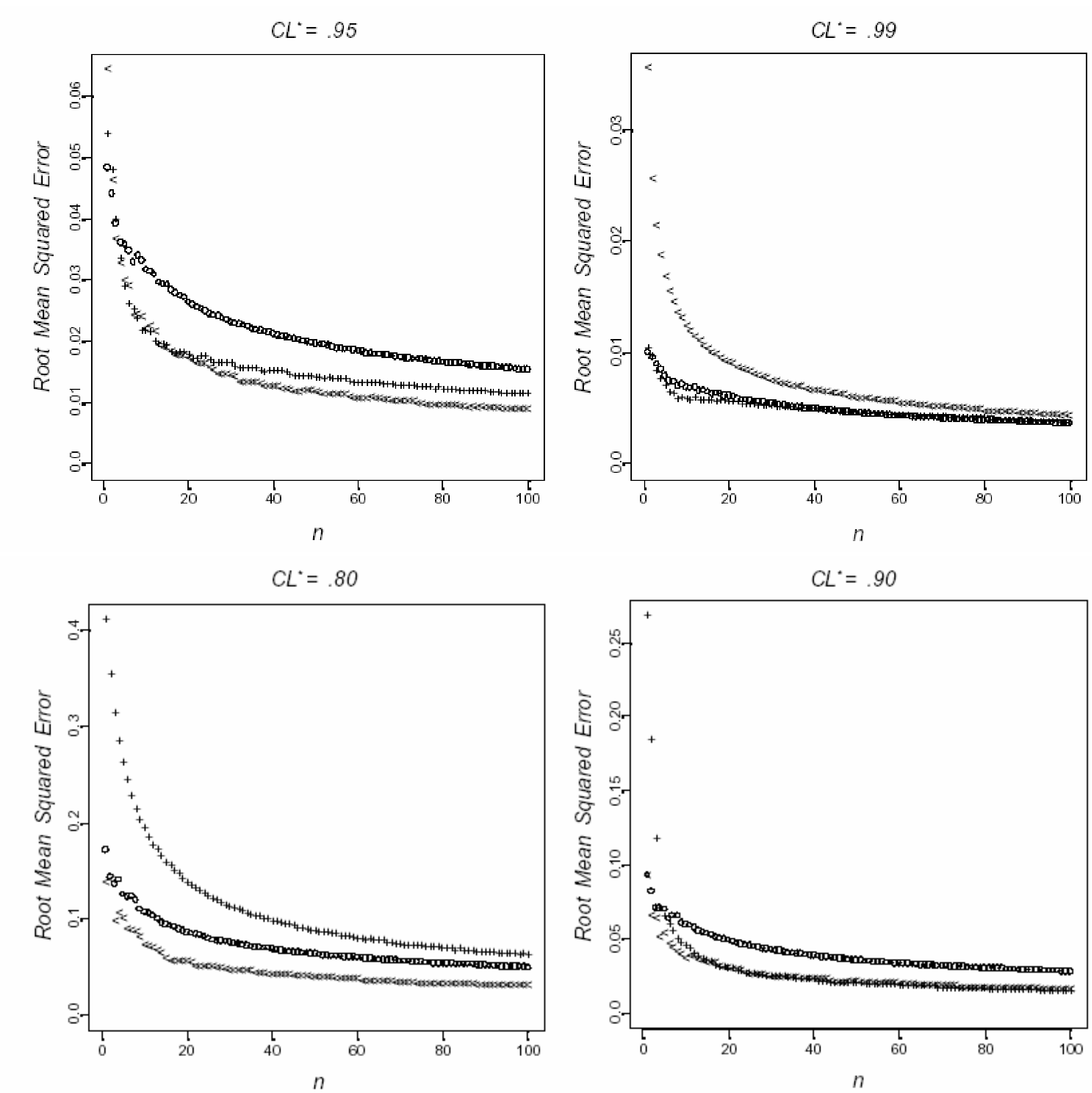


Figure 4. Proportion of Values of p Where Coverage is Closer to Nominal. The scatter is of the proportion of the uniformly distributed values of p for which the adjusted Wald or score method has actual coverage probability closer to the nominal coverage probability than the optimized method for sample sizes of from 1 to 100 with nominal confidence levels of .80, .90, .95, and .99. The adjusted Wald method is indicated by a “o” and the score method by a “+”. The horizontal dotted line is at 50%. At the .80, .90, and .95 nominal confidence levels both the adjusted Wald and Score method tend to have coverage probabilities closer to the nominal for more than half the range of p sample sizes over 20 and this is also true for the score method at a nominal confidence level of .99. For the adjusted Wald at nominal confidence level of .99, and for both methods with sample sizes less than 20, the coverage probability is closer to the nominal than the optimized method for less than half the range of for p.

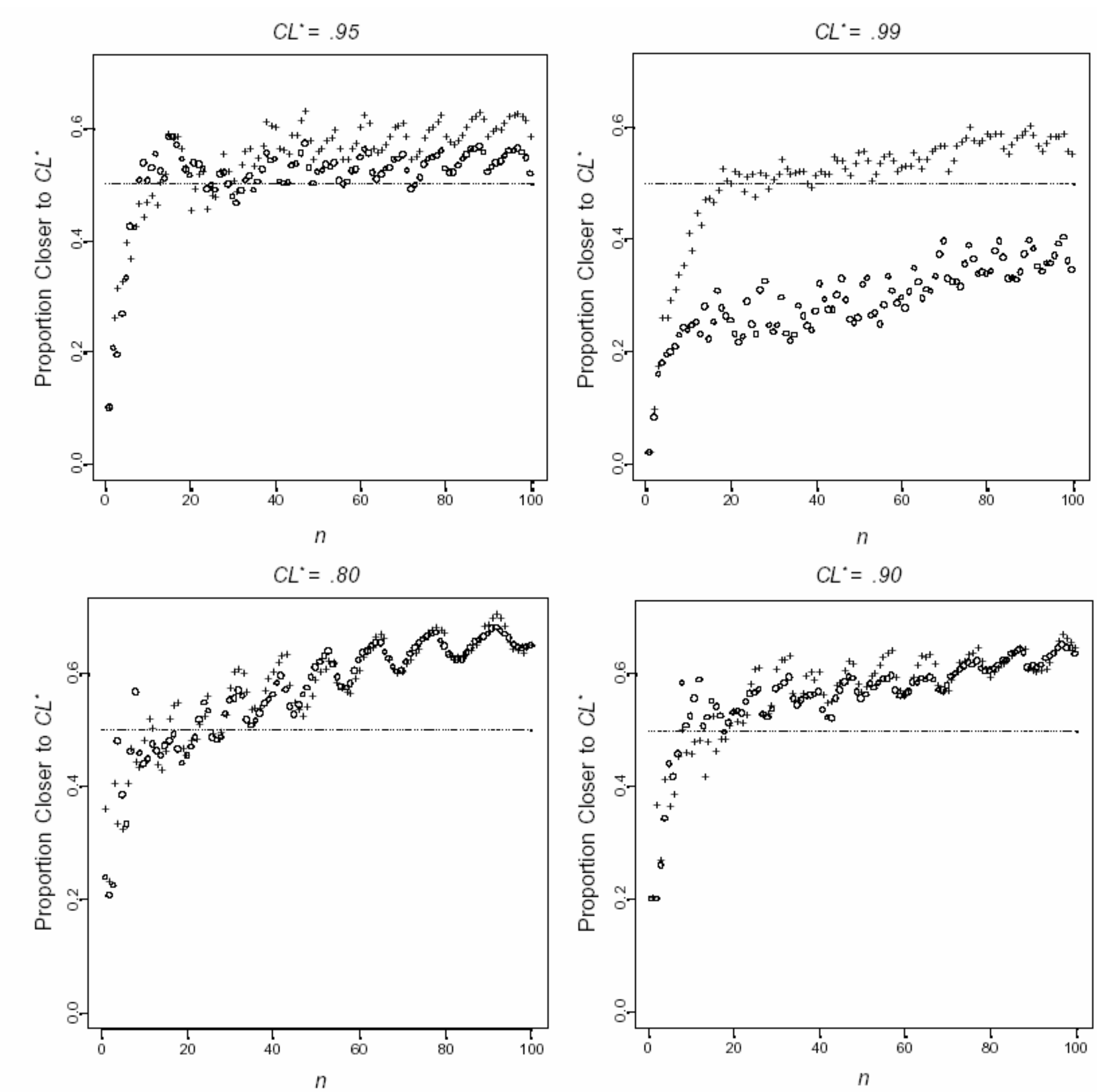


Figure 5. Proportion of p Where Coverage is Less Than the Nominal. The scatter is of the proportion of the uniformly distributed values of p for which a coverage method has actual coverage probability less than the nominal coverage probability for sample sizes of from 1 to 100 with nominal confidence levels of .80, .90, .95, and .99. The optimized interval method is indicated by a “o”, the adjusted Wald method by a “+”, and the score method by a “<”. In general, the optimized interval method has a smaller proportion of the range of p where the actual coverage probability is less than the nominal than the other two methods and this proportion tends to decrease as the sample size increases while it increases for the adjusted Wald and stays at approximately the same level for the score method.

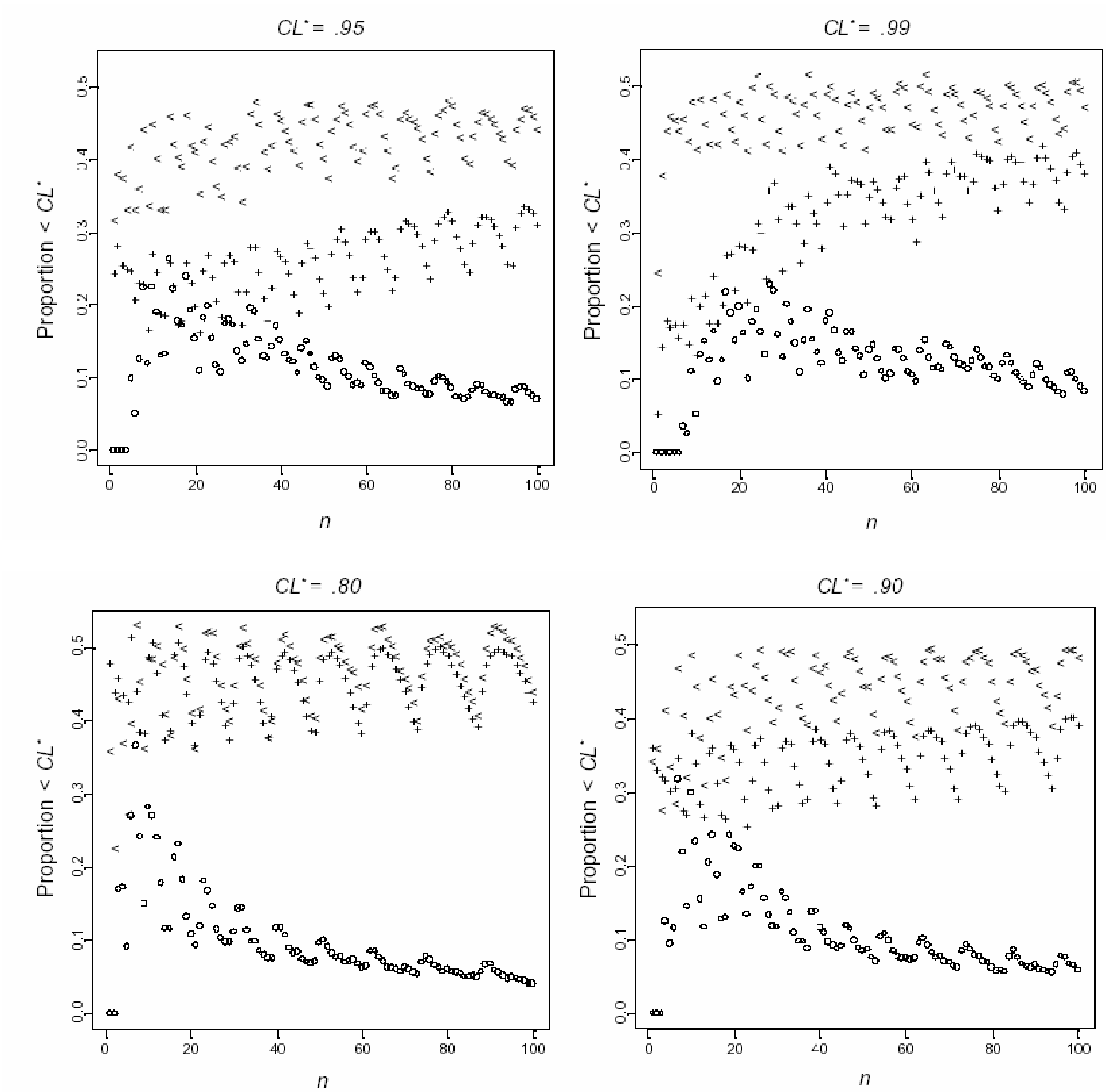
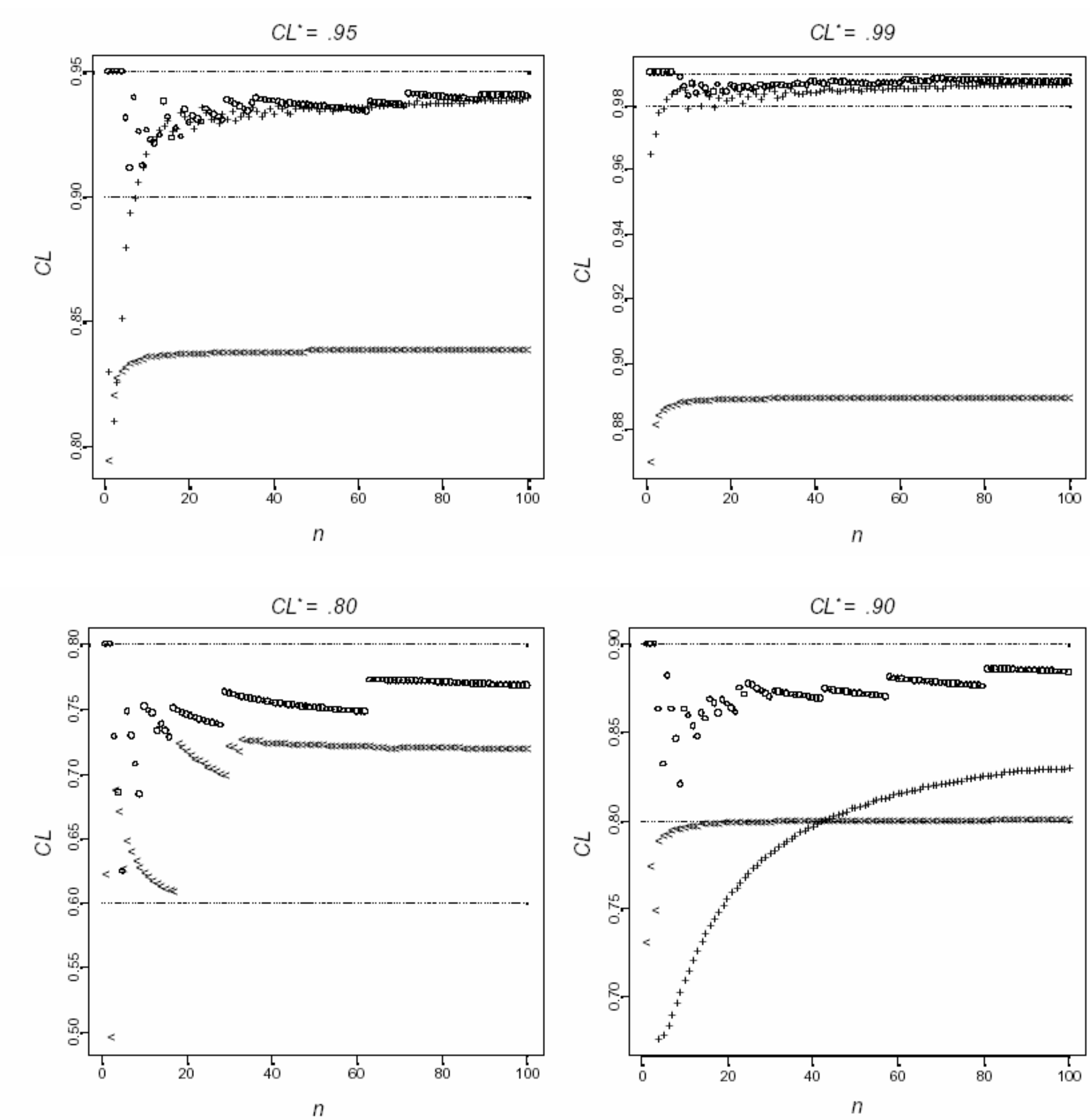


Figure 6. Actual Confidence Levels. The scatter is of the actual confidence levels for three approximate methods for sample sizes of from 1 to 100 with nominal confidence levels of .80, .90, .95, and .99. The optimized interval method is indicated by a “o”, the adjusted Wald method by a “+”, and the score method by a “<”. No actual confidence levels for any sample size are shown for the adjusted Wald method at a nominal confidence level of .80 or for sample sizes less than four at a nominal confidence level of .90. The actual confidence level is zero at all of those points. The upper horizontal dotted line is at the nominal confidence level and the lower dotted line is at the nominal confidence level minus *a*. The actual confidence level for the optimized bound method is always less than nominal level except for very small sample sizes, but it is never less than the nominal level minus *a*. The actual confidence level of the other two methods can be substantially less than the nominal.



Conclusion

The optimized interval method is not an exact method. It should not be used in applications where it is essential that the actual coverage probability be at or above the nominal confidence level across the entire parameter space. For applications where an exact method is not required the optimized method is worth consideration.

Figures 2 – 6 demonstrate that none of the three approximate methods considered in this paper is clearly superior for all of the metrics across all of the sample sizes and nominal confidence levels considered. The investigator needs to determine which metrics are most important and then consult Figures 2 – 6 to determine which method performs best for those metrics at the sample size and nominal confidence level that will be used. If the distance of the actual confidence level from the nominal confidence level and the proportion of the parameter space where coverage falls below the nominal are important considerations then the optimized bound method will often be a good choice.

References

Agresti, A., & Coull, B. A. (1998). Approximate is better than ‘exact’ for interval estimation of binomial proportions, *The American Statistician*, *52*, 119-126.

Berger, R. L., & Coutant, B. W. (2001). Comment on small-sample interval estimation of bernoulli and poission parameters, by D. Wardell. *The American Statistician*, *55*, 85.

Blyth, C. R., & Still, H. A. (1983). Binomial confidence intervals, *Journal of the American Statistical Association*, *78*, 108-116.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*, 404-413.

Wardell, D. G. (1997). Small-Sample interval estimation of bernoulli and poisson parameters. *The American Statistician*, *51*, 321-325.